IT 50700 Measurement and Evaluation of Industry and Technology

Lecture Note

Chapter 6 Descriptive Statistics

Based on the text book: Applied Statistics and Probability for Engineers, 6th Ed, by D. C. Montegomery and G. C. Runger, published by Wiley

6-1 Numerical Summaries of Data
6-2 Stem-and-Leaf Diagrams
6-3 Frequency Distributions and Histograms
6-4 Box Plots
6-5 Time Sequence Plots
6-6 Scatter Diagrams
6-7 Probability Plots

Sample Mean

If the n observations in a sample are denoted by x1, x2, …, xn, the sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_{i=1}^{n} x_i \Big/ n$$

Example 6-1. Calculate the sample mean of pull-off force collected from the prototype engine connectors.

```
% ex6_1.m
% Compute Descriptive Statistics of Pull-Off force of engine connectors.
%
x = [12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1];

sample_mean = mean_s(x)
```

Result:

```
>> ex6_1

sample_mean =     13
```

Function for computing sample mean or sample average

```
% mean.m
% 2016/2/28
% Paul I. Lin
function f = mean_s(x)
N = length(x);
sum = 0.0;
for n = 1: N
    sum = sum + x(n);
end
f = sum/N;
```

MATLAB help commands

>> help function
function Add new function.
New functions may be added to MATLAB's vocabulary if they are expressed in terms of other existing functions. The commands and functions that comprise the new function must be put in a file whose name defines the name of the new function, with a filename extension of '.m'. At the top of the file must be a line that contains the syntax definition for the new function. For example, the existence of a file on disk called STAT.M with:

```
function [mean,stdev] = stat(x)
%STAT Interesting statistics.
n = length(x);
mean = sum(x) / n;
stdev = sqrt(sum((x - mean).^2)/n);
```

defines a new function called STAT that calculates the mean and standard deviation of a vector. The variables within the body of the function are all local variables. See SCRIPT for procedures that work globally on the work-space.

A subfunction that is visible to the other functions in the same file is created by defining a new function with the function keyword after the body of the preceding function or subfunction.
For example, avg is a subfunction within the file STAT.M:

```
function [mean,stdev] = stat(x)
%STAT Interesting statistics.
n = length(x);
mean = avg(x,n);
stdev = sqrt(sum((x-avg(x,n)).^2)/n);

%-------------------------
function mean = avg(x,n)
%AVG subfunction
mean = sum(x)/n;
```

Subfunctions are not visible outside the file where they are defined. You can terminate any function with an END statement but, in most cases, this is optional. END statements are required only in M-files that employ one or more nested functions. Within such an M-file, every function (including primary, nested, private, and subfunctions) must be terminated with an END statement. You can terminate any function type with END, but doing so is not required unless the M-file contains a nested function.

Normally functions return when the end of the function is reached. A RETURN statement can be used to force an early return.

Sample Variance and Sample Standard Deviation
If the n observations in a sample are denoted by x1, x2, …, xn, the sample variance is

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \Big/ (n-1)$$

The **sample standard deviation**, s, is the positive square root of the sample variance.

$$s = \sqrt{s^2} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \Big/ (n-1)}$$

Example 6-2. Sample Variance
Calculate the sample variance of pull-off force collected from the prototype engine connectors.

```
% ex6_2.m
% Compute Descriptive Statistics of Pull-Off force of engine connectors.
%
x = [12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1];

sample_mean = mean_s(x)
Vx = variance_s(x)
```

Compute Sample Variance

```
>> ex6_2
sample_mean =
     13
Vx =
  0.22857
```

Function: Variance – first Version

```
% variance.m
% 2016/2/28
% Paul Lin
function Vx = variance_s(x)
N = length(x);
sum = 0.0;
for n = 1: N
    sum = sum + x(n);
end
mean = sum/N;
s_square = 0.0;
for n = 1: N
    s_square = s_square + (x(n) - mean)^2;
end
Vx = s_square/(N-1);
```

Function: Variance – 2nd Version

```
% variance_s2.m
% 2016/2/28
% Paul Lin
function Vx = variance_s2(x)
N = length(x);
x_avg = mean_s(x);

s_square = 0.0;
for n = 1: N
    s_square = s_square + (x(n) - x_avg)^2;
end
Vx = s_square/(N-1);
```

Compute Sample Variance

```
% ex6_2.m
% Compute Descriptive Statistics of Pull-Off force of engine connectors.
%
x = [12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1];
sample_mean = mean_s(x)

Vx = variance_s(x)

Vx2 = variance_s2(x)
```

Results:

```
>> ex6_2
sample_mean =
      13
Vx =
   0.22857
Vx2 =
   0.22857
```

Example 6-3. Sample Standard Deviation

Calculate the sample standard deviation of pull-off force collected from the prototype engine connectors.

```
% ex6_3.m
% Compute Descriptive Statistics of Pull-Off force of engine connectors.
%
x = [12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1];

standard_dev = sqrt(variance_s(x))
```

How Does the Sample Variance Measure Variability?

For the pull-off force measurement exercise

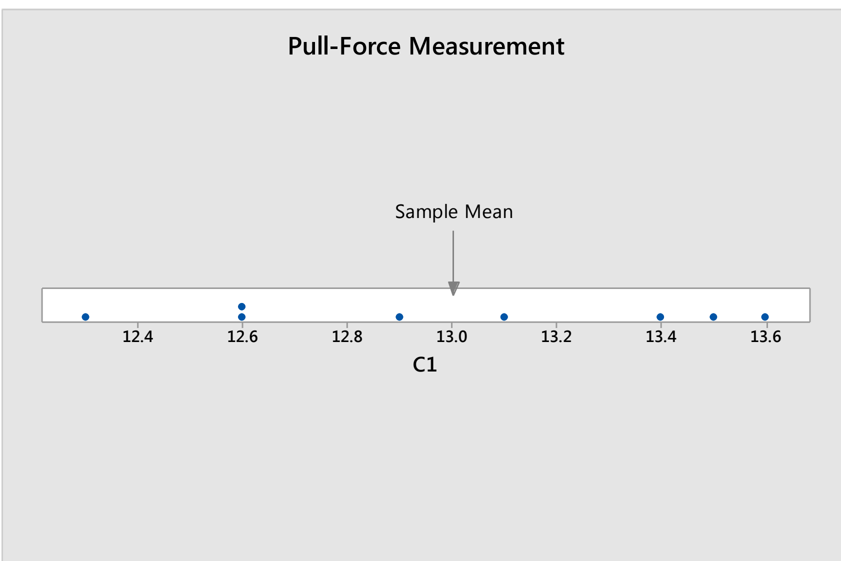
- Sample mean 13 pounds
- Sample standard deviation S = 0.48 pounds
- N-1 = 7 degree of freedom
- Sample Range r = max(xi) – min(xi) = 13.6 – 12.3 = 1.3 pounds

Dot diagram: with deviations $x_i - \bar{x}$

MINITAB Plots:

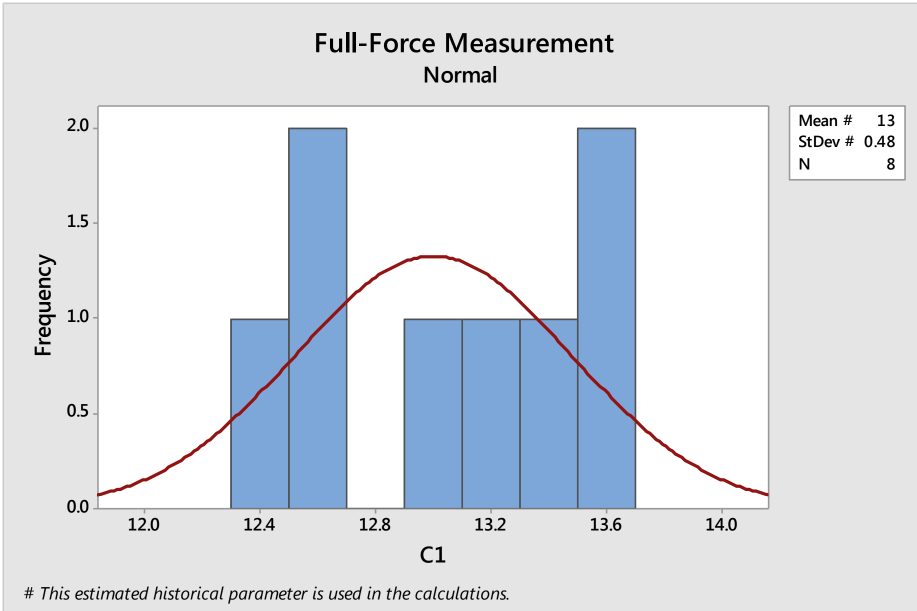
- **Dotplot diagram**
- **Histogram (with data view – normal distribution)**
- **Stem-and-Leaf diagram**
- **Box plot**
- **Probability plot**

Graph => Dotplot => One Y Simple



Histogram Plot

Graph => Histogram
Data View => Distribution: Normal => Historical Parameters: Mean 13, StDev 0.48



Stem-and-Leaf Diagrams

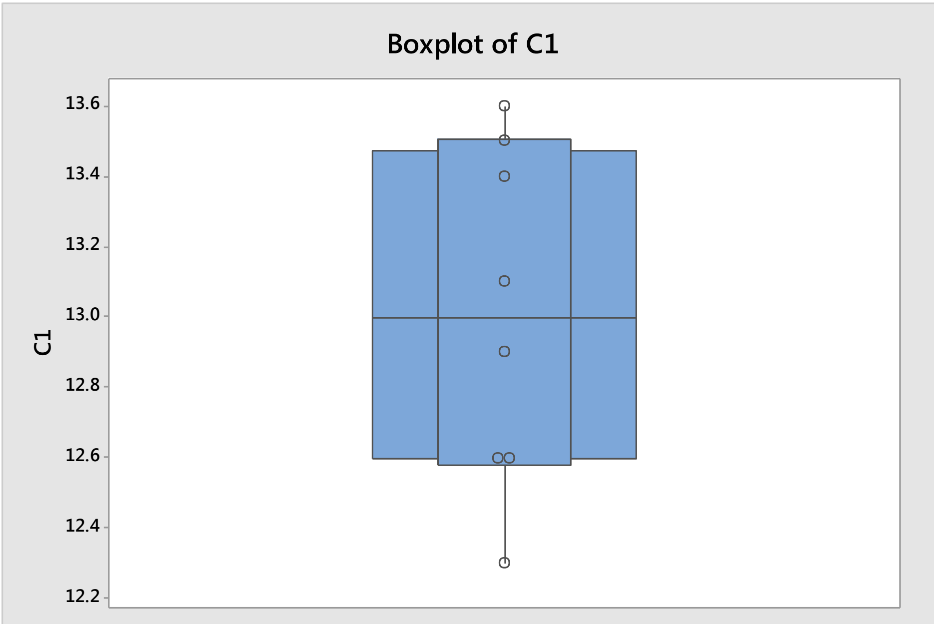
- Graph => Stem-and-Leaf
- X = [12. 6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1]

```
1  12
3  12  66
4  12  9
4  13  1
3  13
3  13  45
1  13  6
```

Box Plot

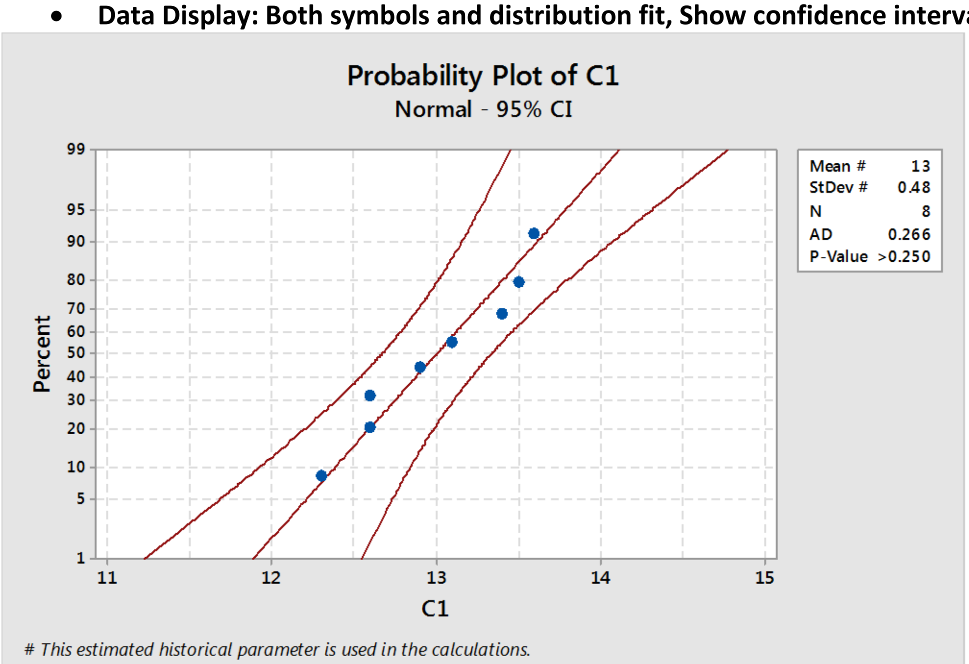
Graph => Box Plot;
Data View:

- Median confidence interval box
- Interquartile range box
- Outlier symbols
- Individual symbol



Probability Plot:

- **Distribution: Normal, Mean = 13, StDev = 0.48**
- **Data Display: Both symbols and distribution fit, Show confidence interval: 95**



6-2 Stem-and-Leaf Diagrams

Example 6-4: Alloy Strength

Stem Number: 7, 8, 9, 10, 11, 12, … 24
Leafs: one digits, frequency count

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Minitab

Graph => Stem-and-Leaf

Stem-and-Leaf Display: C1

```
Stem-and-leaf of C1  N  = 80
Leaf Unit = 1.0

 1    7   6
 2    8   7
 3    9   7
 5    10  15
 8    11  058
 11   12  013
 17   13  133455
 25   14  12356899
 37   15  001344678888
(10)  16  0003357789
 33   17  0112445668
 23   18  0011346
 16   19  034699
 10   20  0178
 6    21  8
 5    22  189
 2    23  7
 1    24  5
```

Median = (40^{th} (160) + 41^{st} (163))/2 = 161.5

Sample Mode = 158 (most frequently occurring data): 4 times(158, 158, 158, 158)

Quartile: divide into 4 quarters

- The first or lower quartile, q1, (approximately 25% of the observation below, and 75% above it)
- The second quartile, q2, (approximately 50% of the observation below its value)
- The third or upper quartile, q3, (approximately 75% of the observation below its value)

Interquartile Range (IQR) = q3 – q1 (a measure of variability)

Minitab

Stat => Basic Statistics => Display Descriptive Statistics

Results for: Table6-2-Column.MTW

Descriptive Statistics: C1

```
Variable   N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3  Maximum
C1        80   0  162.66     3.78  33.77    76.00  143.50  161.50  181.00   245.00

Range
169.00
```

6-3 Frequency Distributions and Histograms

Table 6-2 Compressive Strength (in psi) of 80 Aluminum-Lithium Alloy Specimens

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Minitab

Graph => Histogram

Data View: Fit distribution, Normal, Mean = 162.66, StDev = 33.77

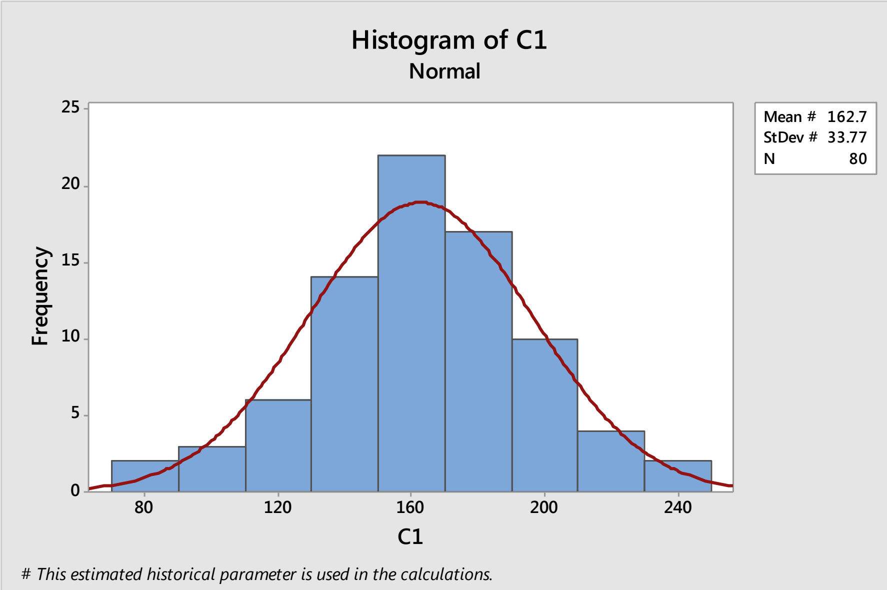
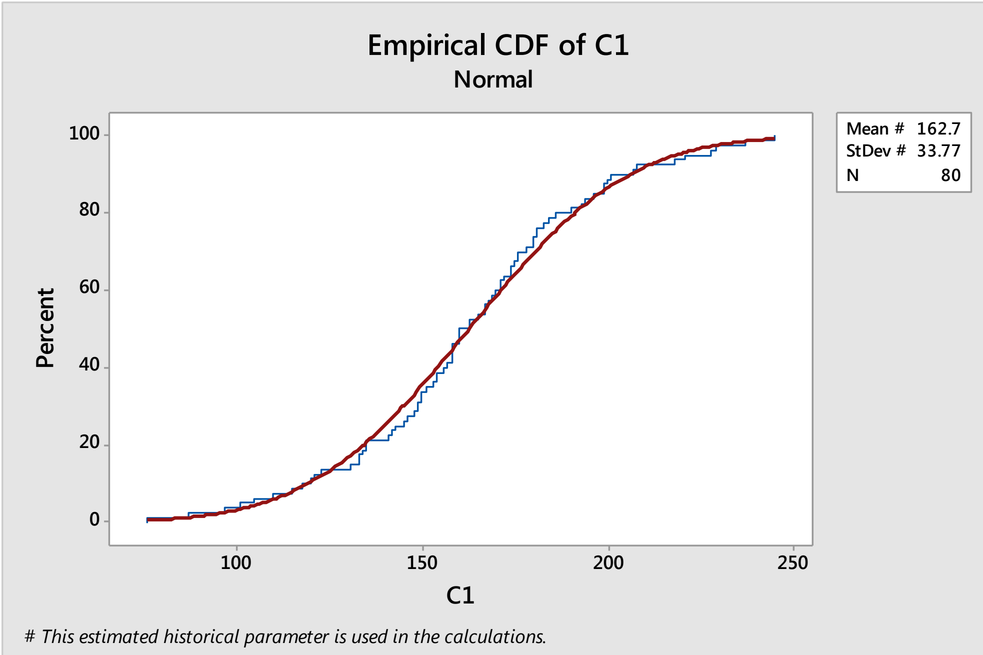


Table 6-4: Frequency Distribution for the Compressive Data in Table 2

Class	70: 90	90:110	110:130	130:150	150:170	170:190	190:210	210:230	230:250
Frequency	2	3	6	14	22	17	10	4	2
Relative Frequency	0.025	0.0375	0.0750	0.1750	0.2750	0.2125	0.1250	0.050	0.025
Cumulative Relative Frequency	0.025	0.0625	0.1375	0.3125	0.5875	0.8000	0.9250	0.9750	1.000

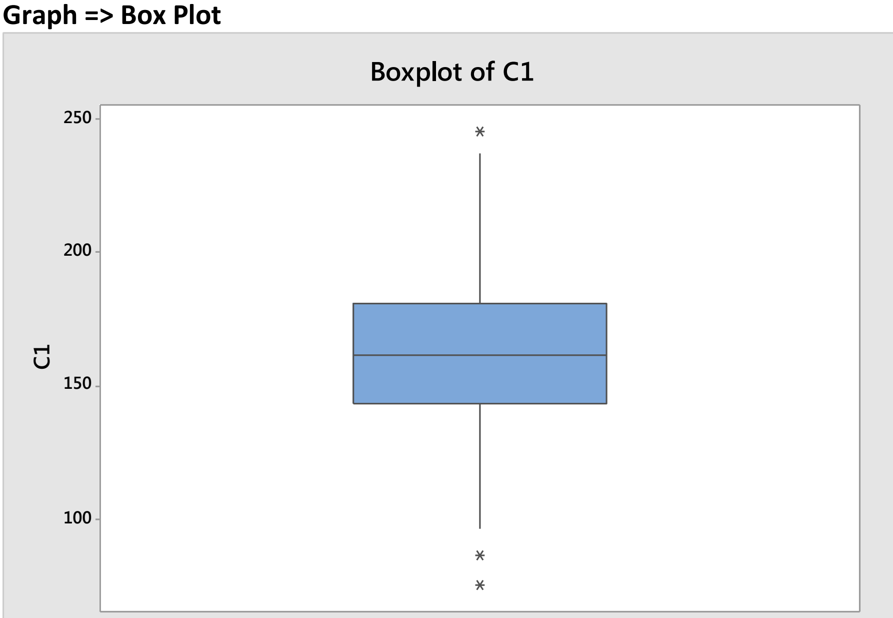
Graph => Empirical CDF (Compare cumulative distribution with a known one)

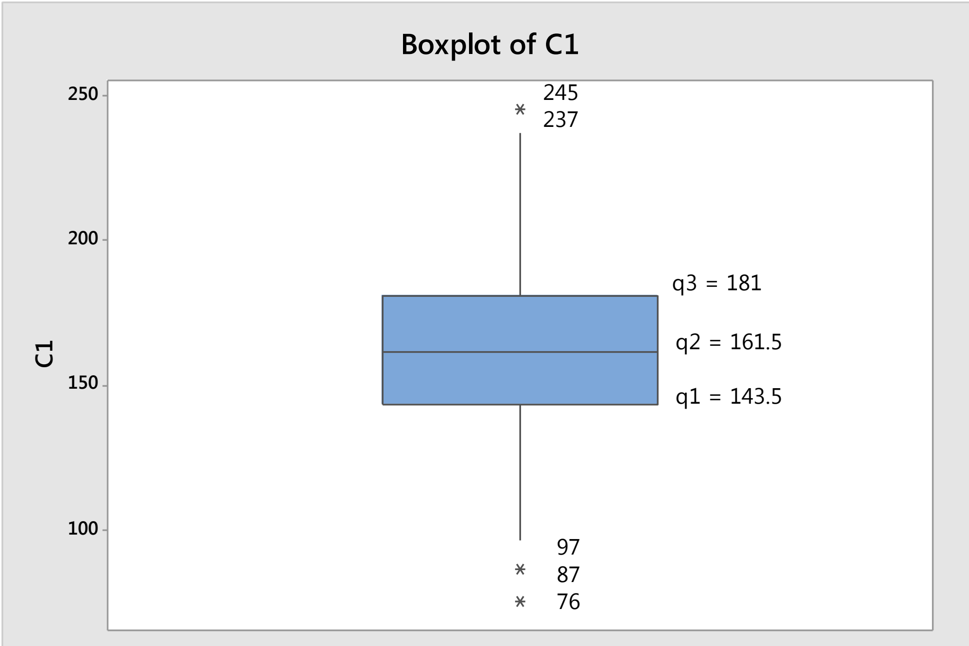


6-4 Box Plots

Minitab

Graph => Box Plot





6-5 Time Sequence Plots